Summary of the 2017 NCGAS User Survey

Thomas G. Doak
Craig A. Stewart
Scott D. Michaels

Indiana University

PTI Technical Report PTI-TR17-005

July 31, 2017

Citation:

Doak, T.G., Stewart, C.A., Michaels, S.D. (2017) "Summary of the 2017 NCGAS User Survey", Indiana University, Bloomington, IN. PTI Technical Report PTI-TR17-005. Retrieved from http://hdl.handle.net/2022/21614

PERVASIVE TECHNOLOGY INSTITUTE
INDIANA UNIVERSITY

RESEARCH TECHNOLOGIES
INDIANA UNIVERSITY
University Information Technology Services
Pervasive Technology Institute

Table of Contents

1. Introduction

Intent of survey: Following our survey of 2 yrs. ago, this was a satisfaction survey, and an attempt to learn more about the activities and needs of our current users. A specific interest was learning more about the funding our users worked from—the last survey didn't do a good job of drawing that out.

2. Methods:

This survey was conducted over March and April of 2017. Users were identified as holders of NCGAS allocations and researchers NCGAS has provided affiliate accounts for. This generated a mailing list of 222, and 56 responses were returned, for a return rate of 25%. Appendix 1 Invitation Letter; Appendix 2 Survey.

3. Results

Who responded: Only 3 respondents were IU-associated, so we did reach our national user group. This is in part because local users are under-identified from our mailing list. 30% had used NCGAS as a graduate student—only two as undergraduates (probably under-identified in the mailing list). ~30% were primary holders of NCGAS allocations; others were on allocations of others, or reported having no affiliation to an allocation (a problem). 50% reported hearing of NCGAS from a colleague (word of mouth), but 20% did report that they learned of us at a conference or workshop, so our outreach is having some success—word of mouth can still originate for a workshop presentation.

Funding sources: We asked what funding sources users had (Fig.1). 60% had NSF funding; thus 40% allocations/accounts are without NSF funding. The 40% may include IU/IUPUI users that don't need NSF funding to get in; some will be doing work consistent with NSF-funded projects and/or generated preliminary results for an NSF proposal. 15% had NIH funding; this can include researchers who also had NSF funding (this could be extracted from the results). A number of other funding agencies played minor roles.

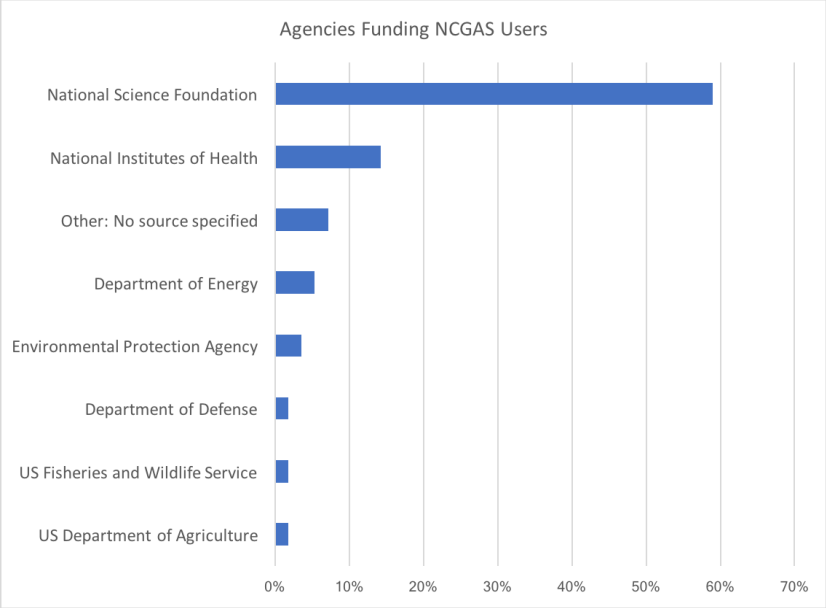

Figure 1

Methods used (data and analysis): Many users (37%) had RNAseq data (Fig.?), but when asked what analysis they used NCGAS for (Fig.?), RNAseq methods did not dominate (RNAseq analysis is split between "metagenomics/transcriptomics" and "*de novo* transcriptomic assembly"). Many users are doing genome assembly and population genomics. Population genomics is a relatively new field, so that is a surprise. 13% of our users are already doing long read assemblies (2nd and 3rd generation sequencing methods; these are not users we hear from much). The major category of analysis was metagenomics, but we are phrased the question as "metagenomics/transcriptomics" by which we meant metatranscriptomics, but this may have been misunderstood.

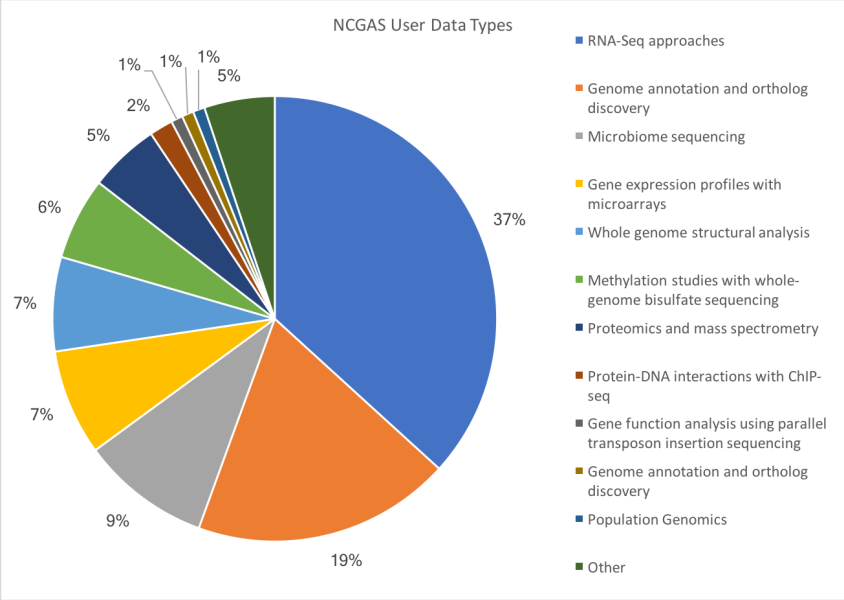

Figure 2

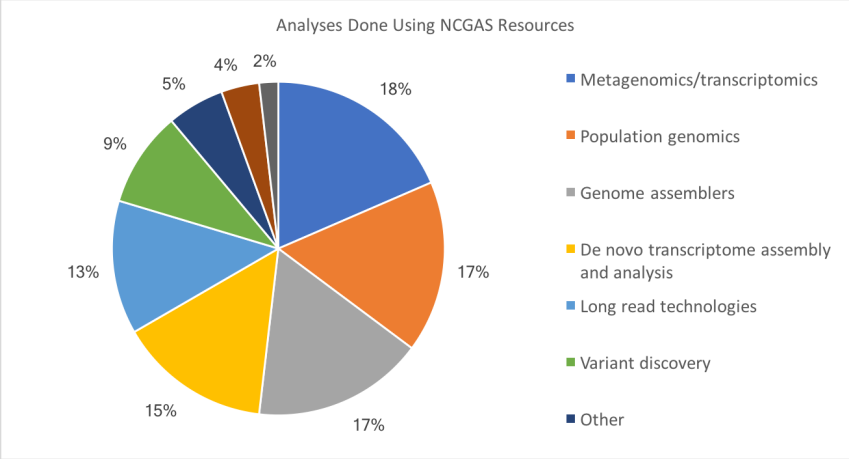

Figure 3

Resources used and needed:

We asked two somewhat similar questions, first, what cyberinfrastructure did the respondents use, NCGAS or anything else (Fig.?). The majority reported using Mason, followed by their own cluster, and then a range of other resources, including Bridges and Amazon. Second, we asked with of NCGAS' services they had used (Fig.?): Mason dominated, followed by help desk/technical support. Browsers was only 5%, given is a new service; Trinity Galaxy was <5%, because its users are not represented in the ailing list.

Finally, we asked them what of the NCGAS services they found useful (Fig.?); the majority of NCGAS's services were found useful by >35% of respondents. Reference data and Galaxy tools were only ~15%; this is consistent with previous polling, which shows that the majority of our users are still using line command, either because they are long-time users, or because they are learning, now that they are confronted with a need for bioinformatics. Given the number of fairly basic linex questions we answer, the latter is at least partly the case.

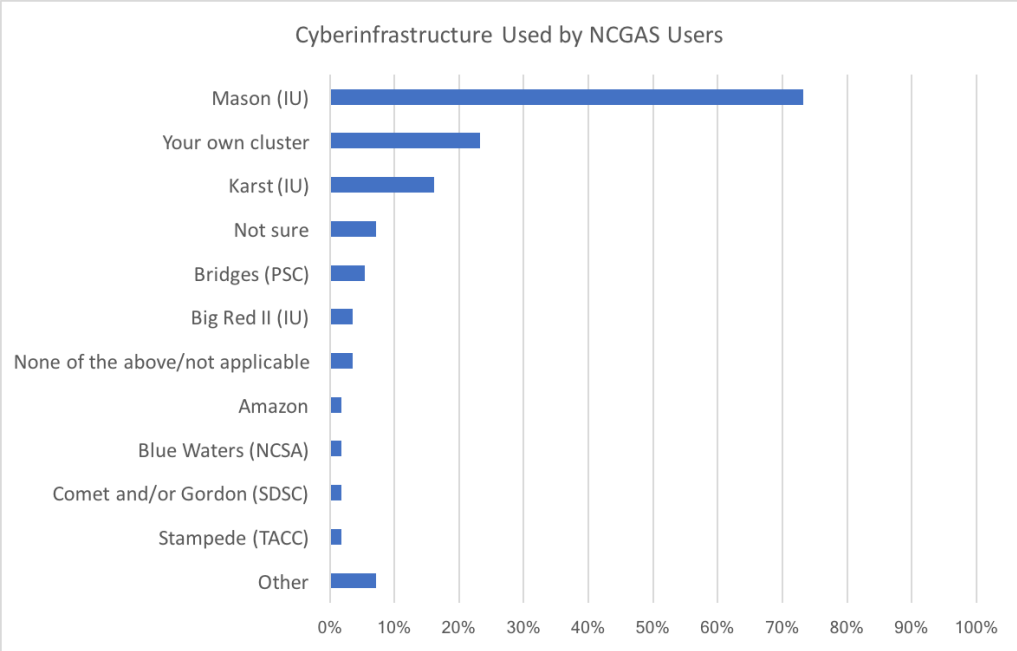

Figure 4

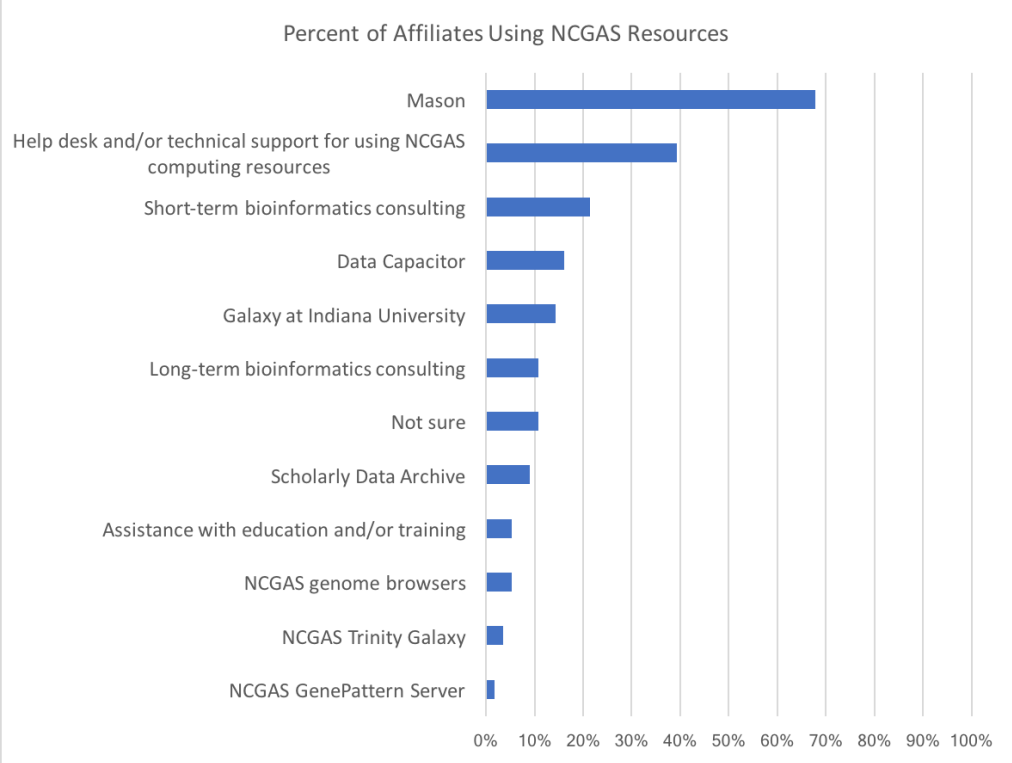

Figure 5

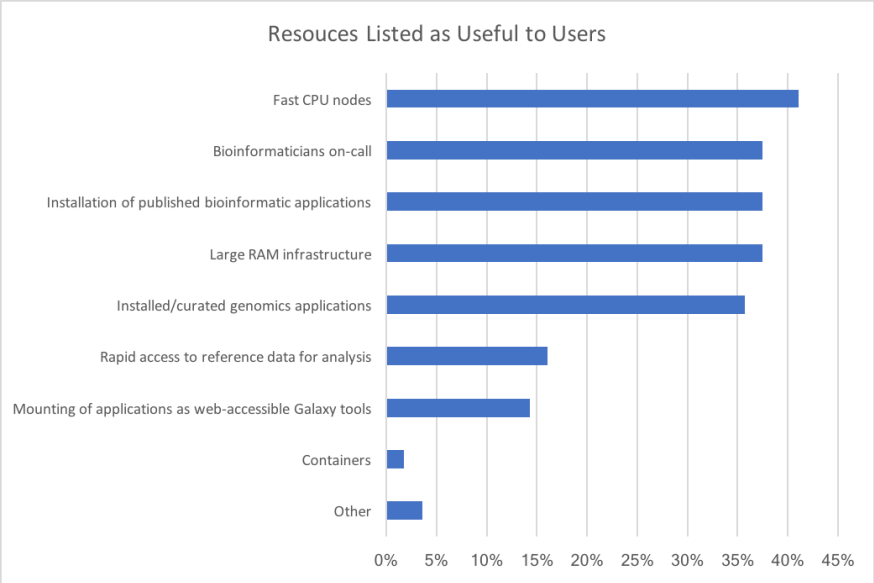

Figure 6

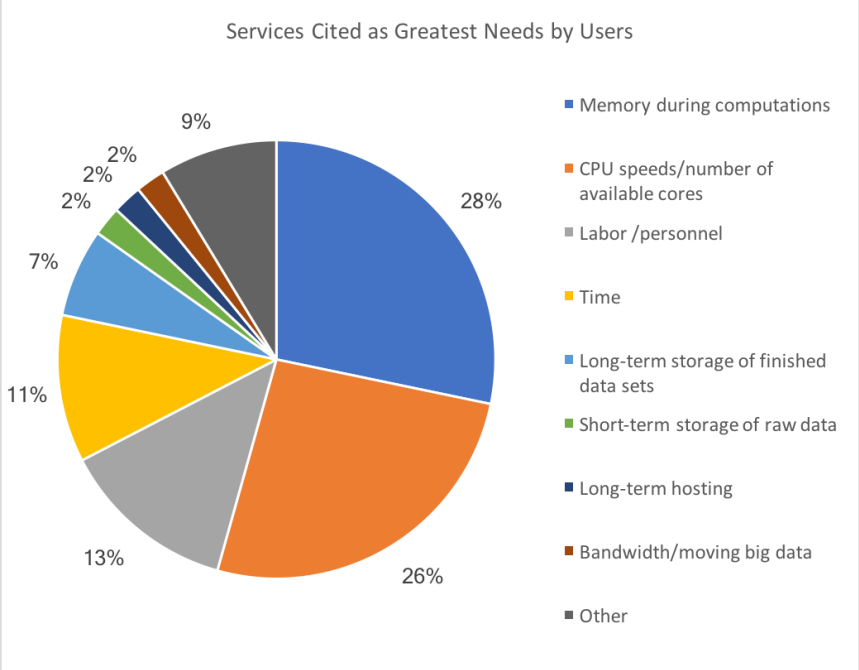

Figure 7

Satisfaction:

Gratifyingly, 60% of respondents reported that we were essential to their research(Fig.?). If true of all 222 users, we're having a very large effect on research. Another 30% said we were helpful, if not essential.

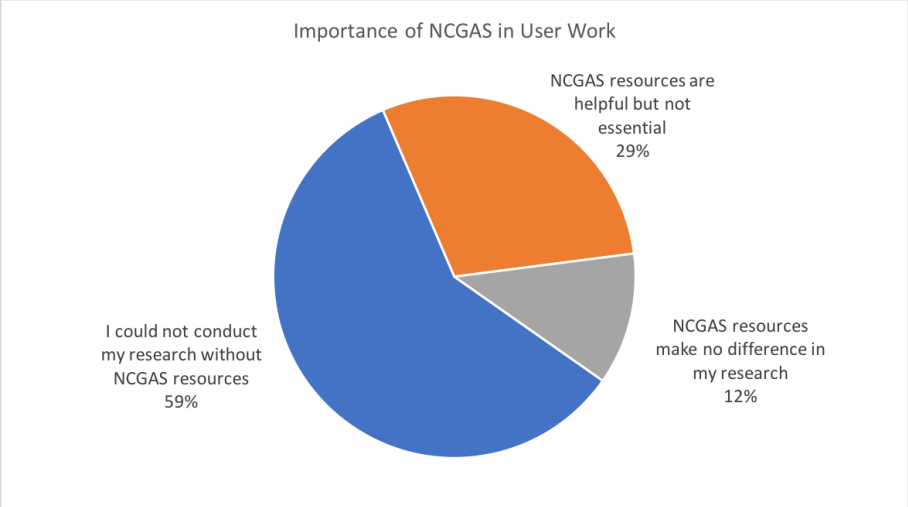

Figure 8

Additional comments from short answers:
In general comments are very complimentary: complementing the above results, access to large memory clusters, access to installed software, direct assistance were all valued. Complaints all point to a lack of staffing, which we aware of: slow answering tickets, not enough on-line materials, generally slow to respond. Clearly this is something that should be improved, but it is hard to see how to do this with current staffing.

4. Conclusions:

We are doing the right things; a big computer is important; we need more personnel.

5. Appendix 1. Invitation Letter

Date: March 30, 2017

To: NCGAS Users

From: Tom Doak, Ph.D.

Subject: NCGAS User Survey – Your feedback is requested

Dear Bioscience Researcher,

As a user of its resources, you likely know the National Center for Genome Analysis Support (NCGAS) at Indiana University is an initiative aiming to provide NSF-funded life scientists with access and support to specialized computational resources on the national cyberinfrastructure. Our goal is to create and sustain a single virtual system that lowers the computational barriers for biologists, bioinformaticians, physician-scientists, life-science students, and anyone, such as yourself, who is conducting research that uses next-generation DNA sequencing.

Toward this goal, we are surveying our users to ensure that the needs of bioscience researchers are at the forefront of our efforts as we look forward to the growth of NCGAS services. As we aim to provide services that best match the current and future needs of the community, especially NSF-funded life scientists, we ask that you take a few moments to share with us your thoughts our services and your experiences in interacting with our Center. The survey should take not more than 10 minutes to complete. Your feedback will be used to improve and expand NCGAS services to the user community and to aid in the decision-making processes related to resource allocation.

Your responses will remain completely confidential. Neither your name nor your organization will be associated with any data you provide or included in any reports. If you have any questions about this survey or how the results will be used, please feel free to contact Julie Wernert, Information Manager, Indiana University, at jwernert@iu.edu, or (812) 856-5517.

Sincerely,

-Tom

Thomas Doak, Ph.D.
Research Scientist and Manager
National Center for Genome Analysis Support (NCGAS.org)
Indiana University Pervasive Technology Institute

Follow this link to the Survey:
${l://SurveyLink?d=Take the survey}

Or copy and paste the URL below into your internet browser:
${l://SurveyURL}

Follow the link to opt out of future emails:
${l://OptOutLink?d=Click here to unsubscribe}

6. Appendix 2. Survey Questions

National Center for Genome Analysis Support User Survey

INFORMED CONSENT:
You are invited to participate in the National Center for Genome Analysis Support (NCGAS) User Survey conducted by principal investigators of the NCGAS project. We ask that you read this statement and ask any questions you may have before agreeing to take part in the survey. This study is administered on behalf of the NCGAS project by the Indiana University Pervasive Technology Institute and is funded, in part, by the National Science Foundation.

STUDY PURPOSE:
The purpose of the NCGAS User Survey is aimed at assessing current and future computational needs for researchers at field and marine stations, as well as the factors informing the adoption and use of cloud-based resources. Survey information will be used to guide NCGAS personnel in (1) assessing and addressing current and future needs, (2) focusing outreach and training efforts, (3) making decisions related to resource provisioning, and (4) informing plans for future expansion of resources and services.

PROCEDURES FOR THE STUDY:
If you agree to take part in the study, you will complete an online survey in which you will not be required to provide any identifying information. The survey will remain confidential, and survey responses will not be associated with any identifying information, even if you choose to or inadvertently disclose such information. You will receive via email an initial letter of invitation, followed by up to three (3) reminder messages. After the initial letter of invitation, only those who have not responded will receive subsequent messages. You will have the opportunity to opt out of all future communications upon receipt of the initial letter of invitation. The survey should not take more than 10 minutes to complete, with an average time for completion in the seven- to eight-minute range.

CONFIDENTIALITY:
Efforts will be made to keep any personal information that you might inadvertently disclose confidential. We cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Your identity will be held in confidence in reports in which the survey results may be published and/or databases in which results may be stored. Organizations that may inspect and/or copy survey records for quality assurance and data analysis include groups such as the study investigator and his/her research associates, the Indiana University Institutional Review Board or its designees, the study sponsor, the National Science Foundation, and (as allowed by law) state or federal agencies, specifically the Office for Human Research Protections (OHRP).

CONTACTS FOR QUESTIONS OR PROBLEMS:
For questions about the study, contact Indiana University Information Manager Julie Wernert at (812) 856-5517 or jwernert@iu.edu. For questions about your rights as a participant or to discuss problems, complaints, or concerns about a research study, to obtain information, or to

offer input, please contact the IU Human Subjects Office at (812) 856-4242 or by email at irb@iu.edu.

VOLUNTARY NATURE OF STUDY:
Taking part in this study is voluntary. You may choose not to take part or may leave the survey at any time. Leaving the survey will not result in any penalty. Your decision whether or not to participate in this survey will not affect your current or future relations with the Indiana University Pervasive Technology Institute, the NCGAS project, or the National Science Foundation. This study was approved by the Indiana University Institutional Review Board on March 29, 2017. Please reference study #1407478943/exempt when inquiring. Do you agree to participate in this survey?

- ❍ I agree (1)
- ❍ I disagree (2)

Condition: I agree Is Selected. Skip To: Please tell us which of the followin....Condition: I disagree Is Selected. Skip To: End of Block.

Please tell us which of the following your research involves. Select all that apply.

- ❑ RNA-Seq approaches (1)
- ❑ Microbiome sequencing (2)
- ❑ Whole genome structural analysis, i.e., nucleosome mapping, high-resolution DNase sensitivity, etc (3)
- ❑ Gene function analysis using parallel transposon insertion sequencing (4)
- ❑ Protein-DNA interactions with ChIP-seq (5)
- ❑ Methylation studies with whole-genome bisulfate sequencing (6)
- ❑ Gene expression profiles with microarrays (7)
- ❑ Proteomics and mass spectrometry (8)
- ❑ Genome annotation and ortholog discovery (9)
- ❑ Other: (10) _____

Which NCGAS services and/or resources have you (or a member of our group) used? Select all that apply.

- ❑ Short-term bioinformatics consulting (1)
- ❑ Long-term bioinformatics consulting (2)
- ❑ Help desk and/or technical support for using NCGAS computing resources (3)
- ❑ Assistance with education and/or training (4)
- ❑ Mason (5)
- ❑ Galaxy at Indiana University (6)
- ❑ Data Capacitor (7)
- ❑ Scholarly Data Archive (8)
- ❑ NCGAS Trinity Galaxy (9)
- ❑ NCGAS GenePattern Server (10)
- ❑ NCGAS genome browsers (11)
- ❑ Not sure (12)
- ❑ None of the above/not applicable (13)

Which of the following computing resources do you use in your work?

- ❑ Mason (IU) (1)
- ❑ Karst (IU) (2)
- ❑ Big Red II (IU) (3)
- ❑ Comet and/or Gordon (SDSC) (4)
- ❑ Bridges (PSC) (5)
- ❑ Stampede (TACC) (6)
- ❑ Blue Waters (NCSA) (7)
- ❑ Jetstream (IU/TACC) (8)
- ❑ Your own cluster (9)
- ❑ Not sure (10)
- ❑ None of the above/not applicable (11)
- ❑ Other: (12) _____

How did you learn about NCGAS? Select all that apply.

- ❑ Conference or workshop (1)
- ❑ Academic journal, paper or website (2)
- ❑ Course lecture or class presentation (3)
- ❑ Press coverage (4)
- ❑ Newsletter, email solicitation, or survey (5)
- ❑ From a colleague (word of mouth) (6)
- ❑ Other: (7) _____

What percentage of your bioinformatics computation is done using NCGAS resources?

_____ Percentage of bioinformatics computation using NCGAS resources (1)

Q4 How essential are NCGAS resources to the execution of your research project(s)?

- ❍ I could not conduct my research without NCGAS resources (1)
- ❍ NCGAS resources are helpful but not essential (2)
- ❍ NCGAS resources make no difference in my research (3)

Condition: NCGAS resources make no dif... Is Selected. Skip To: Please indicate the national resource....Condition: NCGAS resources are helpful... Is Selected. Skip To: Please elaborate on how NCGAS resourc....Condition: I could not conduct my rese... Is Selected. Skip To: Please elaborate on how NCGAS resourc....

Please elaborate on how NCGAS resources are useful to your work? Please do not include any identifying information in your comments.

Please indicate the national resource providers you use in your work. Select all that apply.

- ❑ CyVerse (1)
- ❑ XSEDE (2)
- ❑ NCBI (3)
- ❑ Galaxy (4)
- ❑ The BROAD Institute (5)
- ❑ Center for Computational Biology (CCB) (6)
- ❑ DIAG Computing (7)
- ❑ Open Science Grid (8)
- ❑ CIPRES (9)
- ❑ Your local resources (10)
- ❑ Other: (11) _____

For which of the following application types should NCGAS consider providing greater levels of support? Select all that apply.

- ❍ Genome assemblers (1)
- ❍ Long read technologies (e.g., PacBIO, 10x, Nanopore, etc) (2)
- ❍ De novo transcriptome assembly and analysis (e.g., RAMPAGE, Trinity, etc.) (3)
- ❍ Metagenomics/transcriptomics (4)
- ❍ Proteomics (5)
- ❍ Variant discovery (6)
- ❍ Population genomics (7)
- ❍ Other: (8) _____

Which of the following needs/resources would be of the greatest benefit to your research?

- ❍ Short-term storage of raw data (1)
- ❍ Long-term storage of finished data sets (2)
- ❍ Memory during computations (3)
- ❍ CPU speeds/number of available cores (4)
- ❍ Time (5)
- ❍ Labor /personnel (6)
- ❍ Bandwidth/moving big data (7)
- ❍ Long-term hosting (8)
- ❍ Access to reference data (9)
- ❍ Other: (10) _____

Which of the following resources would be helpful to your research program? Select all that apply.

- ❑ Bioinformaticians on-call (1)
- ❑ Large RAM infrastructure (2)
- ❑ Fast CPU nodes (3)
- ❑ Installed/curated genomics applications (4)
- ❑ Installation of published bioinformatic applications (5)
- ❑ Mounting of applications as web-accessible Galaxy tools (6)
- ❑ Rapid access to reference data for analysis (7)
- ❑ Other: (8) _____

Do you have access to a genome browser for your research?

- ❍ Yes (1)
- ❍ No (2)

Would support for genome browsers be helpful to your research program and/or community?

- ❍ Yes (1)
- ❍ No (2)

If you have a preferred genome browser application (e.g., GBrowse, JBrowse, GBase, etc.), please list itbelow:

On a 1-5 scale, with "1" being "very dissatisfied" and "5" being "very satisfied", please rate your satisfaction with the following NCGAS services. If you have no basis for evaluation, please select not applicable.

	Very dissatisfied (1)	Somewhat dissatisfied (2)	Neither satisfied or dissatisfied (3)	Somewhat satisfied (4)	Very satisfied (5)	Not applicable/never used (6)
Genome analysis/bioinformatics consulting (1)	❍	❍	❍	❍	❍	❍
Use of software workflow systems deployed by NCGAS (e.g., Galaxy, GenePattern, JBrowse, etc.) (2)	❍	❍	❍	❍	❍	❍
Use of the Mason cluster (3)	❍	❍	❍	❍	❍	❍
Educational materials (4)	❍	❍	❍	❍	❍	❍
Talks and guest lectures (5)	❍	❍	❍	❍	❍	❍
Response time (6)	❍	❍	❍	❍	❍	❍
Overalls satisfaction with NCGAS (7)	❍	❍	❍	❍	❍	❍

What do you like about NCGAS resources? Please do not include any identifying information in your comments.

How can we improve our support of your research needs? Please do not include any identifying information in your comments. You are always free to contact us at ncgas@iu.edu for project specific comments and requests.

To help us better understand your responses please answer the following demographicquestions. To ensure your survey responses remains confidential, please do not include any identifying information in your comments. If you have questions or concerns and would like to be contacted by NCGAS staff, please contact us at ncgas@iu.edu.

Have you used NCGAS services as an undergraduate or two-year student?
- ❍ Yes (1)
- ❍ No (2)

Have you used NCGAS services as a graduate or PhD student?
- ❍ Yes (1)
- ❍ No (2)

Are you a member of the Indiana University community?
- ❍ Yes (1)
- ❍ No (2)

From which of the following agencies do you have funding? Select all that apply.
- ❑ National Science Foundation (1)
- ❑ National Institutes of Health (2)
- ❑ Department of Energy (3)
- ❑ US Department of Agriculture (4)
- ❑ Department of Defense (5)
- ❑ Environmental Protection Agency (6)
- ❑ Other: (7) _____

Have you been funded as a principal investigator or co-principal investigator on a National Science Foundation grant award?
- ❍ Yes (1)
- ❍ No (2)

Are you funded by the NSF for projects that use NCGAS resources?
- ❍ Yes (1)
- ❍ No (2)

Are the activities for which you have used NCGAS resources in an area of research funded by the National Science Foundation?
- ❍ Yes (1)
- ❍ No (2)

Please best describe your status as a user of NCGAS services and resources.
- ❍ I am the owner of an NCGAS allocation (1)
- ❍ I am on someone else's allocation (2)
- ❍ I am not associated with an NCGAS allocation (3)

If you have published any work that used or acknowledges NCGAS services or resources, please send us a separate email to ncgas@iu.edu with information about your publication. This is the best way to document our contributions, and is extremely important to our continued funding.